BRILL

Multisensory Research 26 (2013) 581–592

brill.com/msr

Short Note

Multisensory Encoding Improves Auditory Recognition

Zachary D. Moran [1,*], **Peter Bachman** [2], **Phillip Pham** [1], **Seong Hah Cho** [1], **Tyrone D. Cannon** [1,2,3] **and Ladan Shams** [1,4]

[1] Department of Psychology, University of California, Los Angeles, USA
[2] Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, USA
[3] Department of Psychology, Yale University, New Haven, CT, USA
[4] Department of Bioengineering, University of California, Los Angeles, USA

Received 19 April 2013; accepted 16 December 2013

Abstract

Recent studies have challenged the long-held belief that recognition is unfailingly degraded by contextual differences between study and test items. In these studies, recognition of pictures presented in silence was better when during study or initial exposure the images were accompanied by a semantically congruent sound rather than silence. In the present study, we sought to examine the generalization of this phenomenon to auditory recognition and found a significant improvement in the recognition of auditory items when coupled with a congruent picture. We discuss these findings within the framework of the redintegration hypothesis of memory retrieval as well as Bayesian inference and learning.

Keywords

Multisensory, recognition

1. Introduction

Memory research has almost invariably shown that recognition performance is optimal when the stimulus context at recognition testing (e.g., an image) is identical to that presented at initial exposure (i.e., the study item; e.g., Hirshman and Bjork, 1988). Indeed, even minor changes from encoding to query context have been shown to be deleterious to recognition (Biederman, 2000). For example, if the background of an object is changed from uniform white to

---

[*] To whom correspondence should be addressed. E-mail: zdmoran@ucla.edu

DOI:10.1163/22134808-00002436

a rich scene or *vice versa* between study and test, it can cause a decrease in recognition performance (Hayes *et al.*, 2007).

On the other hand, a few recent studies have shown that recognition of visual images of an object presented in silence is superior for objects that were initially presented together with a semantically related sound, compared to those that were studied in the absence of sound (Lehmann and Murray, 2005; Murray *et al.*, 2004, 2005; Thelen *et al.*, 2012). Here, not only does a change in the context (pairing with sound *vs.* silence) not deteriorate recognition, it facilitates performance. In these studies, participants were presented with a stream of interleaved new and repeated line drawings and instructed to indicate whether each was ‘old’ or ‘new’. A subset of the initial study items was coupled with a corresponding sound. Image recognition was better for items coupled initially with sound, despite the absence of that sound at test.

This effect occurs only when the sound is semantically congruent to the visual memoranda (Lehmann and Murray, 2005; Thelen *et al.*, 2012). Similarly, a benefit in recognition of audio-visual pairs learned during segregated study and test phases has been shown to be limited to congruent pairs as well (Naghavi *et al.*, 2011). However, Nyberg *et al.* (2000) showed an absence of multisensory facilitation in a study in which participants were asked to actively recall contextual information about picture memoranda (i.e., whether it was or was not originally presented with a sound) also within the context of separate study and test phases. Thus, the facilitatory effect of multisensory encoding may be unique to recognition memory and only under conditions of semantic congruence of the individual sensory components.

Other researchers (e.g., Nyberg *et al.*, 2000; Wheeler *et al.*, 2000) have posited that multisensory encoding of memories may be underpinned by a distributed network of perceptual/memory traces such that the activation of one node is sufficient for activation of an entire object representation in whole (Haxby *et al.*, 2001; Rolls *et al.*, 1997). This type of distributed object recognition process has been termed ‘redintegration’ in cognitive psychology (Hamilton, 1859; Horiwitz and Prytulak, 1969; Nyberg *et al.*, 2000; Shams and Seitz, 2008; Tulving and Madigan, 1970). If redintegration represents a general mechanism for memory encoding and retrieval, then one would expect that this effect should generalize to recognition of sounds and other stimuli. It has been previously shown that learning to identify voices can be facilitated by pairing of vocal audio with the corresponding facial image (Von Kriegstein and Giraud, 2006) or video (Sheffert and Olsen, 2004). Here we aimed to examine whether *recognition* (i.e., declarative memory) of *object sounds* can benefit from pairing of object sounds with corresponding object images. To address this question, we compared the recognition of sounds (in the absence of any images) that during study were presented together with their corresponding image *versus* those that during study were presented without any images. Further, in a follow-up experiment we examined the role of semantic congruency between sounds and images during study on the subsequent recognition of unisensory sounds. In line with previous findings on visual recognition (Lehmann and Murray, 2005; Thelen *et al.*, 2012), we hypothesized that the multisensory encoding benefit would be confined to semantically congruent auditory–visual pairs. This experiment also served the function of controlling for the effects of mere alerting effect of images presented during encoding beyond that of multisensory integration *per se*.

2. Experiment 1

2.1. Aims

In experiment 1, we sought to investigate whether auditory–visual encoding can facilitate subsequent recognition of sounds.

2.2. Methods

2.2.1. Ethics Statement

This study was approved by the UCLA Institutional Review Board (IRB) and experimenters obtained written informed consent from all participants.

2.2.2. Participants

Eleven participants from the UCLA Psychology subject pool engaged in this experiment, eight of whom were women (72.7%). Mean age was 19.64 years (SD = 1.03). They received course credit for participation. All participants reported normal or corrected-to-normal vision and normal hearing.

2.2.3. Materials and Stimuli

All sounds were played on two Harmon/Kardon speakers situated on either side of the computer screen, while visual stimuli were presented on a Dell Inspiron laptop computer with a 15″ monitor. One hundred and sixty eight sounds were drawn from multiple online databases and edited to be 500 ms in length, each with equal maximum audio intensity, at a sampling rate of 44 100 Hz using the freeware program Audacity (http://audacity.sourceforge.net). They were presented at a comfortable but easily audible volume which was constant across all conditions and participants. Picture stimuli consisted of 168 colored sketches drawn both from the database of Snodgrass and Vanderwart-Like Objects (SVLO, Rossion and Pourtois, 2004) and from an online database (www.office.microsoft.com), modified to be stylistically similar to the SVLO pictures using Adobe Photoshop (www.adobe.com).

2.2.4. Procedure

We adapted the paradigm used by Murray and colleagues (i.e., a serial old–new judgment task) for use with auditory memoranda. That is, rather than using images as primary memoranda, we presented participants with a series

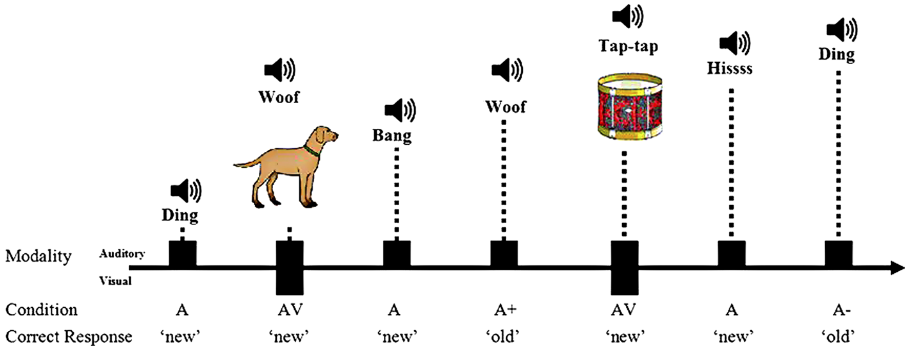

Figure 1. Experimental task (figure and task adapted from Lehmann and Murray, 2005). This figure is published in colour in the online version.

of sounds and the observer's task was to judge whether each sound was old or new (see Fig. 1 for a visual depiction of experimental procedure). Participants listened to a series of sounds made by common objects, presented in a pseudorandom order, each of which was presented twice within the series. They were asked to indicate by button press, as quickly and accurately as possible, whether each sound was novel ('new') or a repeat ('old') of an earlier sound having been informed that 'some sounds may be repeated'.

At initial presentation, half of the sounds were presented alone (the A condition) while the other half were presented as a sound plus a congruent picture (e.g., a bird chirping coupled with the image of a bird — AV condition). Repeated presentation of sounds always occurred in the absence of any picture. Adapting the notation used by Murray *et al.* (2004), we will refer to the repeat presentations of sounds that were initially presented without pictures as condition $A-$ and those that were repeats of sounds initially presented with a picture as condition $A+$. Because all repeated stimuli were sounds alone, the visual stimuli were considered task-irrelevant and participants were instructed not to report on anything image-related.

Stimulus presentation was programmed in Matlab (http://www.mathworks.com) using the Psychophysics Toolbox, Version 3 (http://psychtoolbox.org). Sounds and pictures were both presented for 500 ms, followed by a random inter-trial interval of between 1.7 and 2 s during which a fixation cross was presented. Sounds presented in the absence of pictures maintained the presence of fixation cross. Pseudorandomization of stimulus order was accomplished through computer programming at the outset of the experiment by randomly assigning a sound to a given stimulus order position while ensuring that both the mean and median number of pictures between an initial and repeated exposure was 17 across both conditions of the experiment. The 336 trials were grouped into four blocks of 84, each spanning approximately four minutes and presented in a randomized order. Thus, while stimulus order within blocks was consistent across participants, the ordering of blocks themselves varied randomly from person to person. While no sound was repeated across blocks, participants were not informed of this and, in the event that they asked, they were told that it was possible. Participants were granted a very short break (no longer than one minute) between blocks.

2.3. Results

In order to examine the hypothesis that recognition of auditory stimuli is superior when encoded as part of a multisensory pair, we used a paired t-test comparing mean hit-rate between A+ and A− recognition. We found a significant difference between these conditions, $t(\text{df} = 10) = 4.64$, $p < 0.01$ [*mean* (SD): A+ = 0.85 (0.09), A− = 0.76 (0.11)], corresponding to a large effect size of (Cohen's d) 1.47 standard deviations (with correction for dependence between paired measures, Morris and DeShon, 2002; see Fig. 2). Similarly, mean reaction time was significantly lower for A+ stimuli relative to A−, $t(\text{df} = 10) = 2.94$, $p < 0.05$ [mean (SD): A+ = 0.98 s (0.15), A− = 1.00 s (0.16); see Fig. 3].

When prompted for qualitative input regarding strategies used in performing this task, no participant reported an awareness of the fact that all repeated stimuli were sounds alone (i.e., a sound is more likely to be 'new' if paired with a picture). However, because encoding hit rates for multisensory (AV) were better than encoding hit rates for unisensory counterparts (A) $t(\text{df} = 10) = 4.56$, $p < 0.01$, we conducted the following analyses to examine whether the above effect was a consequence of encoding differences. The difference in encoding hit rates between AV and A conditions (i.e., $\text{AV}_{\text{hit rate}} - \text{A}_{\text{hit rate}}$) was not significantly correlated with hit rates in

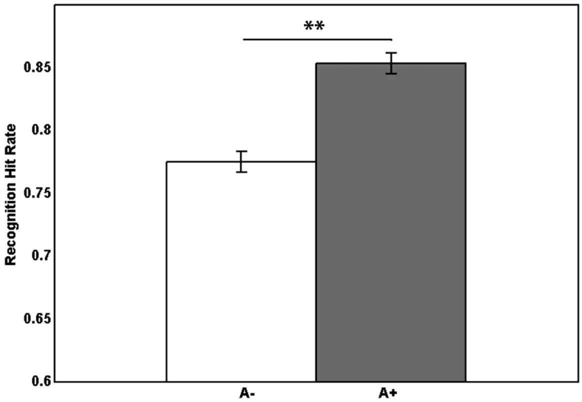

Figure 2. Recognition hit rates for experiment 1. $N = 11$, error bars indicate within-subject standard error (see Barakat *et al.*, 2013 for a computational example). ** = $p < 0.01$.

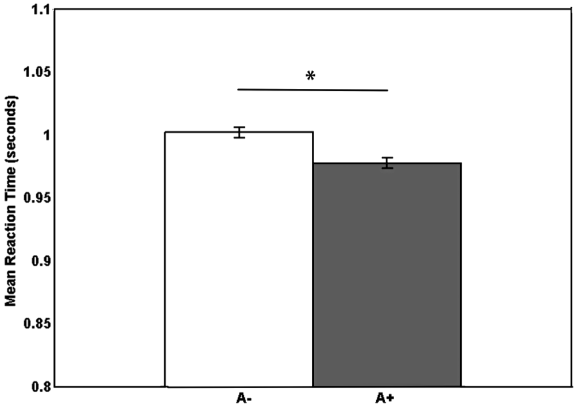

Figure 3. Mean reaction times for experiment 1. $N = 11$, error bars indicate within-subject standard error. * = $p < 0.05$.

either unisensory (A−) or multisensory (A+) recognition conditions (A−: $r = 0.380$, $p = 0.249$; A+: $r = 0.281$, $p = 0.403$). Similarly, the addition of difference in encoding hit rates (i.e., $\mathrm{AV}_{\text{hit rate}} - \mathrm{A}_{\text{hit rate}}$) as a covariate to the repeated-measures comparison of means between hit rate in unisensory (A−) and multisensory (A+) recognition conditions did not impact the significance of the finding, $F(1, 9) = 11.22$, $p < 0.01$.

2.4. Discussion

Findings from experiment 1 generalize the multisensory encoding benefit, which had previously been shown only in recognition of images, to the recognition of sounds. The size of this effect was quite large, showing nearly 1.5 standard deviations improvement in the recognition of sounds encoded with pictures *vs.* without.

3. Experiment 2

3.1. Aims

In this experiment, we investigated whether semantic congruency between the sound and image during study is required for facilitation of auditory recognition. If the observed facilitation is due to an alerting effect of images during study, then a similar facilitation should be observed with incongruent images, as they are equally salient and thus, equally alerting.

3.2. Methods

3.2.1. Participants

Thirty six participants from the UCLA Psychology subject pool engaged in the experiment, 28 of whom were women (77.8%). Mean age was 20.56 years

(SD = 1.52). They were pseudorandomly assigned to one of the two conditions: congruent condition and incongruent condition, yielding 18 participants per condition. They received course credit for participation. As before, all participants reported normal or corrected-to-normal vision and normal hearing.

3.2.2. Materials and Stimuli

While experimental stimuli matched those of experiment 1 in regards to the images and sounds that were used, participants completed the task on a different computer system consisting of an iMac 6 coupled with a 19″ monitor and Roland DM-10 speakers. As in experiment 1, sound intensities were fixed across conditions and participants, here set specifically to an average peak of 70.3 dB.

3.2.3. Procedure

The procedure for experiment 2 was identical to that of experiment 1 except that there was an additional condition in which AV sounds were coupled with pseudorandomly-assigned semantically-incongruent images. To avoid carry-over effects, a between-group design was used, with one group presented with congruent AV trials and another group presented with incongruent AV trials. All other trials were identical between the two groups.

3.3. Results

All data (i.e., those from both experiment 1 and experiment 2) were examined using a 2 (encoding modality: auditory or audio-visual) × 2 (audio-visual congruence condition, congruent or incongruent) repeated-measures ANOVA, whereby the former variable was within-subject and the latter was between subject. This revealed significant main effects of both encoding modality, $F(1, 45) = 13.55$, $p < 0.05$ (i.e., sounds encoded with pictures showed better recognition), and congruence, $F(1, 45) = 11.69$, $p < 0.01$ (i.e., participants engaging in the AV-congruent condition showed better overall recognition), as well as qualification of those effects by a trend-level encoding × congruence interaction, $F(1, 45) = 3.21$, $p = 0.08$.

Examining data from experiment 2 alone revealed a significant increase in the mean hit-rate of recognition for sounds which were encoded as part of a congruent multisensory pair thus replicating the results of experiment 1, $t(\text{df} = 17) = 2.61$, $p < 0.05$, [mean (SD): $A+ = 0.74$ (0.12), $A- = 0.69$ (0.12) — see left panel of Fig. 4]. In contrast, we found no significant difference in the mean-hit rate for recognition of sounds which were encoded as part of an incongruent multisensory pair, $t(\text{df} = 17) = 1.28$, $p > 0.05$, [mean (SD): $A+ = 0.66$ (0.09), $A- = 0.63$ (0.08) — see right panel of Fig. 4].

No significant differences were found in mean reaction time as a function of either congruence *vs.* incongruence ($p > 0.05$) or unisensory *vs.* multisensory encoding [$p > 0.05$; grand mean (SD) = 0.92 s (0.13), see Fig. 5].

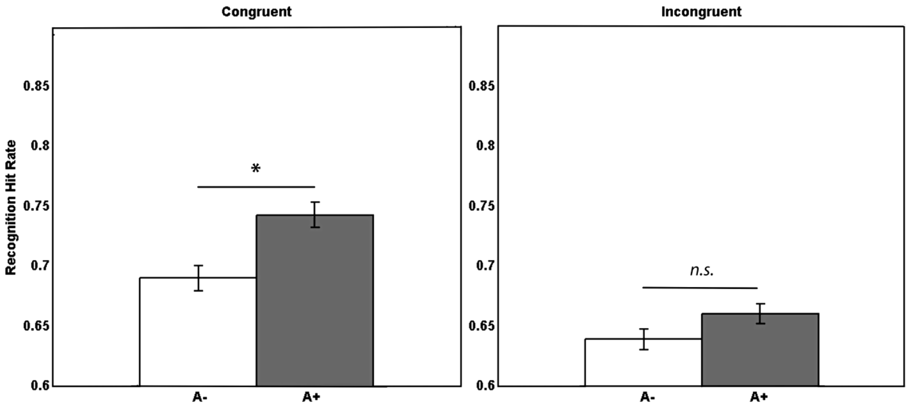

Figure 4. Left: Recognition hit rates for congruent condition of experiment 2. $N = 18$, error bars indicate within-subject standard error. $^* = p < 0.05$. Right: Recognition hit rates for incongruent condition of experiment 2. $N = 18$, error bars indicate within-subject standard error. n.s. = not significant, $p > 0.1$.

3.4. Discussion

Findings from experiment 2 are in line with previous findings in the visual recognition literature in showing specificity of multisensory benefit to encoding in which stimulus-pairs are semantically congruent (Lehmann and Murray, 2005; Thelen *et al.*, 2012). While sounds that were paired with a semantically congruent picture were better recognized than those that were encoded without a picture, such benefit was not observed for sounds that were paired with semantically incongruent pictures.

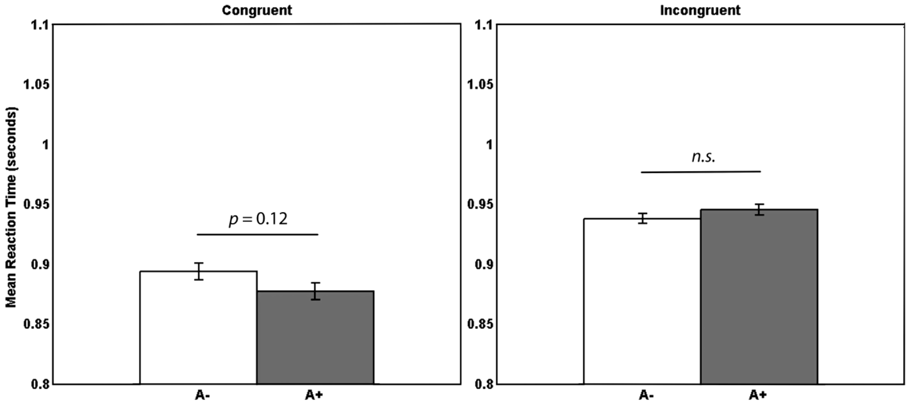

Figure 5. Left: Mean reaction times for congruent condition of experiment 2. $N = 18$, error bars indicate within-subject standard error. Right: Mean reaction times for incongruent condition of experiment 2. $N = 18$, error bars indicate within-subject standard error. n.s. = not significant, $p > 0.1$.

4. General Discussion

The primary finding of this study is that recognition of auditory stimuli is facilitated by encoding which includes a semantically congruent image. Thus, the current findings generalize the multisensory encoding benefit, which had previously been shown only in recognition of images, to the recognition of sounds.

Consistent with what has been proposed by other authors (e.g., Nyberg *et al.*, 2000; Wheeler *et al.*, 2000), we interpret these results to be suggestive of a mechanism of memory retrieval akin to redintegration, which we take to be a general property of memory retrieval rather than a property only of visual recognition. For an object that was encoded using an auditory stimulus alone, the test stimulus will only activate the auditory representations. However, for an object that is encoded using both sound and an image, the auditory test stimulus may activate both the auditory and visual memory traces. Thus, in case of the latter, retrieval would operate on a richer and more informative representation, resulting in a higher accuracy. The idea that activation of a cortically-distributed multisensory memory trace occurs *via* re-activation of one of its sensory elements has been supported by a growing body of neuroimaging research. In each of these studies, unisensory recognition items which were previously learned in the context of a multisensory set activated cortical regions corresponding both to the sensory modality of the recognition item *and* that of the modality of the item with which it was originally paired (Gottfried *et al.*, 2004; Nyberg *et al.*, 2000; Wheeler *et al.*, 2000). Thus, in our study, activation of a larger, audio-visual cortical network by re-exposure to sounds originally matched with pictures may underlie their increased recognition in this behavioral paradigm. We suggest as a possible mechanism for this redintegration-like memory facilitation the predictive coding model of object representation (Friston, 2005; Rao and Ballard, 1999), whereby, once the association between two object features — here, auditory and visual features — has been established, the presentation of partial data — here, object sound — gives rise to the prediction of the missing data — here, object image.

An alternative mechanism that could underlie the observed benefit of multisensory encoding is a learning mechanism by which unisensory representations are updated by multisensory experience (Jacobs and Shams, 2010; Shams *et al.*, 2011). In this Bayesian model (Jacobs and Shams, 2010), the inference about a property s of an object (e.g., identity of an object) is based on the posterior probability of the property given the sensory input, $p(s|I)$. When the nervous system is presented with two sensory inputs, I_A, and I_V, for the same object, the multisensory inference of the object, $p(s|I_\mathrm{A}, I_\mathrm{V})$, serves as a teaching signal for each unisensory inference, $p(s|I_\mathrm{A})$ and $p(s|I_\mathrm{V})$. Therefore, the auditory representation formed by the nervous system in presence of visual input will be superior to that formed in the absence of visual input. This superior representation will then give rise to superior subsequent unisensory recognition.

These two mechanisms are theoretically inter-related. In the latter framework, the repeated exposure to paired congruent auditory–visual representations will eventually render $p(s|I_{\mathrm{A}})$ and $p(s|I_{\mathrm{V}})$ equivalent to $p(s|I_{\mathrm{A}}, I_{\mathrm{V}})$. In other words, at limit, the exposure to one of the stimuli (visual or auditory) will result in hallucination of the missing stimulus and hence the phenomenon of redintegration.

The Bayesian learning model is a computational model and therefore cannot make predictions about the underlying neural mechanisms. However, some predictions about behavioral phenomena can be derived and tested. For example, this theory would predict a larger cross-modal facilitation of recognition for an impoverished stimulus that is accompanied by a rich/diagnostic cross-modal stimulus during encoding compared to that of a rich stimulus accompanied by an impoverished cross-modal stimulus. Specifically, the recognition of a sound which is ambiguous/non-diagnostic and not easy to recognize (e.g., the sound of a squirrel) would benefit more from pairing with the corresponding image (of a squirrel) compared to recognition of a dog bark which is distinct and easy to recognize. Similarly, if images are richer and easier to recognize than sounds, the recognition of images would benefit less from bisensory encoding than the recognition of sounds.

Another set of predictions has to do with semantic congruency between the paired sounds and images. Assuming the causal inference model of multisensory processing (Kording *et al.*, 2007; Shams *et al.*, 2005; Wozny *et al.*, 2010), the largest cross-modal facilitation should be obtained when the image and the sound are highly congruent (e.g., sound and image of a barking dog), as this would result in perception of a common cause for the two and integration of the two signals which will in turn lead to the updating of unisensory posterior by the bisensory posterior. If the image and sound are highly incongruent (e.g., sound of a dog bark paired with the image of a flower), then the brain would likely infer independent causes. Therefore the two stimuli would not get integrated and there would be no facilitation, and more importantly, no deterioration in recognition (i.e., recognition of $\mathrm{A}+$ will be no worse than that of $\mathrm{A}-$, and similarly for $\mathrm{V}+$ and $\mathrm{V}-$). However, if the two stimuli are only moderately incongruent (e.g., sound of a dog bark and the image of a roaring lion) the nervous system may either partially integrate the two stimuli all the time (if a model averaging strategy is employed; Wozny *et al.*, 2010) or occasionally integrate the two stimuli (if a model selection or probability matching strategy is employed; Wozny *et al.*, 2010). In either case, the partial integration of the incongruent stimuli would result in deterioration (as opposed to facilitation) of recognition on average; i.e., recognition of $\mathrm{A}+$ will be worse than recognition of A− (and similarly for V+ and V−). Future research will need to test these predictions and shed light on mechanisms underlying the benefit of multisensory encoding in unisensory recognition.

Acknowledgements

The authors would like to thank Micah Murray for consultation regarding the implementation of his experimental paradigm, and Brandon Barakat, Mostafa Majidpour, Brian Odegaard, Megan Peters, and Majed Samad for their helpful discussion of study results. LS was supported by NSF grant 1057969.

References

Barakat, B. K., Seitz, A. R. and Shams, L. (2013). The effect of statistical learning on internal stimulus representations: predictable items are enhanced even when not predicted, *Cognition* **129**, 205–211.

Biederman, I. (2000). Recognizing depth-rotated objects: a review of recent research and theory, *Spat. Vis.* **13**, 241–253.

Friston, K. (2005). A theory of cortical responses, *Phil. Trans. R. Soc.* **360**, 815–836.

Gottfried, J. A., Smith, A. P. R., Rugg, M. D. and Dolan, R. J. (2004). Remembrance of odors past: human olfactory cortex in cross-modal recognition memory, *Neuron* **42**, 687–695.

Hamilton, W. (1859). *Lectures on Metaphysics and Logic*, Vol. I. Gould and Lincoln, Boston, MA, USA.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex, *Science* **293**, 2425–2430.

Hirshman, E. and Bjork, R. A. (1988). The generation effect: support for a two-factor theory, *J. Exp. Psychol. Learn. Mem. Cogn.* **14**, 484–494.

Horowitz, L. M. and Prytulak, L. S. (1969). Redintegrative memory, *Psychol. Rev.* **76**, 519–532.

Jacobs, R. A. and Shams, L. (2010). Visual learning in multisensory environments, *Top. Cogn. Sci.* **2**, 217–225.

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLOS One* **2**, e943.

Lehmann, S. and Murray, M. M. (2005). The role of multisensory memories in unisensory object discrimination, *Cogn. Brain Res.* **24**, 326–334.

Morris, S. B. and DeShon, R. P. (2002). Combining effect sizes estimates in meta-anlysis with repeated measures and independent-groups designs, *Psychol. Meth.* **7**, 105–125.

Murray, M. M., Foxe, J. J. and Wylie, G. R. (2005). The brain uses single-trial multisensory memories to discriminate without awareness, *NeuroImage* **27**, 473–478.

Murray, M. M., Michel, C. M., de Peralta, R. G., Ortigue, S., Brunet, D., Andino, S. G. and Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging, *NeuroImage* **21**, 125–135.

Naghavi, H. R., Eriksson, J., Larsson, A. and Nyberg, L. (2011). Coritical regions underlying successful encoding of semantically congruent and incongruent associations between common auditory and visual objects, *Neurosci. Lett.* **505**, 191–195.

Nyberg, L., Habib, R., McIntosh, A. R. and Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval, *Proc. Natl Acad. Sci. U.S.A.* **97**, 11120–11124.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nat. Neurosci.* **2**, 79–87.

Rolls, E. T., Treves, A. and Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex, *Exp. Brain Res.* **114**, 149–162.

Rossion, B. and Pourtois, G. (2004). Revisiting Snodgrass and Vadnerwart's object pictorial set: the role of surface detail in basic-level object recognition, *Perception* **33**, 217–236.

Shams, L., Ma, W. J. and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept, *Neuroreport* **16**, 1923–1927.

Shams, L. and Seitz, A. R. (2008). Benefits of multisensory learning, *Trends Cogn. Sci.* **12**, 411–417.

Shams, L., Wozny, D. R., Kim, R. and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing, *Front. Psychol.* **2**, 264.

Sheffert, S. M. and Olson, E. (2004). Audiovisual speech facilitates voice learning, *Percept. Psychophys.* **66**, 352–362.

Thelen, A., Cappe, C. and Murray, M. M. (2012). Electrical neuroimaging of memory discrimination based on single-trial multisensory learning, *NeuroImage* **62**, 1478–1488.

Tulving, E. and Madigan, S. A. (1970). Memory and verbal learning, *Annu. Rev. Psychol.* **21**, 437–487.

Von Kriegstein, K. and Giraud, A. (2006). Implicit multisensory associations influence voice recognition, *PLoS Biol.* **4**, 1809–1820.

Wheeler, M. E., Petersen, S. E. and Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex, *Proc. Natl Acad. Sci.* **97**, 11125–11129.

Wozny, D. R., Beierholm, U. R. and Shams, L. (2010). Probability matching as a computational strategy used in perception, *PLOS Comp. Biol.* **6**, e1000871.